CLAIMS

What is claimed is:

1. A method for identifying biological markers in a set of n biological measurements for each of p observations, wherein $n > p$ and each observation is associated with a clinical endpoint, each biological marker comprising at most k measurements, wherein $k < p$, said method comprising:
   a) reducing said set of n measurements to a set of m candidate measurements; and
   b) selecting at least two biological markers from said set of m candidate measurements, wherein values of each biological marker predict said clinical endpoints.

2. The method of claim 1, wherein said clinical endpoints correspond to clinical classes.

3. The method of claim 1, wherein said clinical endpoints correspond to a continuous response variable.

4. The method of claim 1, wherein $n > 10p$.

5. The method of claim 1, wherein $k < p/5$.

6. The method of claim 1, wherein step (a) comprises performing a correlation analysis.

7. The method of claim 6, wherein said correlation analysis comprises a correlation-based cluster analysis.

8. The method of claim 7, wherein said correlation-based cluster analysis comprises a correlation-based hierarchical cluster analysis.

9. The method of claim 6, wherein said correlation analysis is performed in part in dependence on a user-selected correlation threshold.

10. The method of claim 6, wherein said correlation analysis is performed in part in dependence on a user-selected value of m.

11. The method of claim 1, wherein step (a) comprises performing a differential significance analysis.

12. The method of claim 11, wherein said differential significance analysis is performed in part in dependence on a user-selected significance threshold.

13. The method of claim 1, wherein said n measurements have different sources.

14. The method of claim 1, further comprising ranking said selected biological markers.

15. The method of claim 14, wherein said biological markers are ranked in dependence on an accuracy of predicting said clinical endpoints.

16. The method of claim 1, wherein said biological markers are selected from all possible subsets of at most k measurements of said set of m measurements.

17. The method of claim 16, wherein said biological markers are selected by evaluating each of said possible subsets.

18. The method of claim 17, wherein said possible subsets are evaluated in parallel.

19. The method of claim 1, wherein step (b) comprises simulated annealing.

20. The method of claim 1, wherein k is a user-selected value.

21. The method of claim 1, wherein k is selected in dependence on a desired computation time.

22. The method of claim 1, wherein m is selected in dependence on a desired computation time.

23. The method of claim 1, further comprising performing a market-basket analysis of said selected biological markers.

24. A method for identifying a biological marker in a set of n biological measurements for each of p observations, wherein $n > p$ and each observation is associated with a clinical endpoint, each biological marker comprising at most k measurements, wherein $k < p$, said method comprising:

a) reducing said set of n measurements to a set of m candidate measurements; and

b) using simulated annealing, selecting a biological marker from said set of m candidate measurements, wherein values of said biological marker predict said clinical endpoints.

25. The method of claim 24, wherein $n > 10p$.

26. The method of claim 24, wherein $k < p/5$.

27. The method of claim 24, wherein step (a) comprises performing a correlation analysis.

28. The method of claim 27, wherein said correlation analysis comprises a correlation-based cluster analysis.

29. The method of claim 28, wherein said correlation-based cluster analysis comprises a correlation-based hierarchical cluster analysis.

30. The method of claim 27, wherein said correlation analysis is performed in part in dependence on a user-selected correlation threshold.

31. The method of claim 27, wherein said correlation analysis is performed in part in dependence on a user-selected value of m.

32. The method of claim 24, wherein step (a) comprises performing a differential significance analysis.

33. The method of claim 32, wherein said differential significance analysis is performed in part in dependence on a user-selected significance threshold.

34. The method of claim 24, wherein said n measurements have different sources.

35. The method of claim 24, wherein k is a user-selected value.

36. The method of claim 24, wherein k is selected in dependence on a desired computation time.

37. The method of claim 24, wherein m is selected in dependence on a desired computation time.

38. The method of claim 24, further comprising performing a market-basket analysis on said selected biological markers.

39. A method for identifying at least one biological marker in a set of n biological measurements for each of p observations, wherein $n > 10p$ and each observation is associated with a clinical endpoint, each biological marker comprising at most k measurements, wherein $k < p$, said method comprising:

a) reducing said set of n measurements to a set of m candidate measurements; and

b) selecting at least one biological marker from said set of m candidate measurements, wherein values of each biological marker predict said clinical endpoints.

40. A program storage device accessible by a processor, tangibly embodying a program of instructions executable by said processor to perform method steps for a biological marker identification method, wherein said method identifies biological markers in a set of n biological measurements for each of p observations, wherein $n > p$ and each observation is associated with a clinical endpoint, each biological marker comprising at most k measurements, wherein $k < p$, said method steps comprising:

a) reducing said set of n measurements to a set of m candidate measurements; and

b) selecting at least two biological markers from said set of m candidate measurements, wherein values of each biological marker predict said clinical endpoints.

41. A program storage device accessible by a processor, tangibly embodying a program of instructions executable by said processor to perform method steps for a

biological marker identification method, wherein said method identifies a biological marker in a set of n biological measurements for each of p observations, wherein $n > p$ and each observation is associated with a clinical endpoint, each biological marker comprising at most k measurements, wherein $k < p$, said method steps comprising:

a) reducing said set of n measurements to a set of m candidate measurements; and

b) using simulated annealing, selecting a biological marker from said set of m candidate measurements, wherein values of said biological marker predict said clinical endpoints.

42. A program storage device accessible by a processor, tangibly embodying a program of instructions executable by said processor to perform method steps for a biological marker identification method, wherein said method identifies at least one biological marker in a set of n biological measurements for each of p observations, wherein $n > 10p$ and each observation is associated with a clinical endpoint, each biological marker comprising at most k measurements, wherein $k < p$, said method steps comprising:

a) reducing said set of n measurements to a set of m candidate measurements; and

b) selecting at least one biological marker from said set of m candidate measurements, wherein values of each biological marker predict said clinical endpoints.